

# BIG DATA ANALYSIS & ITS CHALLENGES

Miss. Priyanka Mishra

Computer Science & Engineering Department, FET, Rama University, Kanpur, U.P, India

\*\*\*

**Abstract** - This literature review paper encapsulates the state-of-the-art research on big data analytics. Because of the exchange of massive amount of data everyday and the need for better data-based decision, businesses nowadays are rebound for better ways to systematic manage, and sharpen these huge datasets. The determined part is to achieve insights from the data and to make use of these refinements. Since big data is primarily unstructured and from prevalent sources, advanced business analytics techniques have to be mandatory to analyze the data. In addition, divergent business analytics and big data concepts can be used to get new insights. Based on these insights, decisions are made which can lead to ruthless assets for an organization. This paper can offer an summary of business analytics and massive information ideas as well as their impacts on competitive benefits of firms. We have a propensity to contribute to theory by providing an summary of huge information and business analytics from a general and technical eventual augment by an summary of fields for future analysis.

**Key Words:** Business Analytics, Big data, Literature Review.

## 1. INTRODUCTION

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to draw out valuable information and grasp, perform detailed analyses, as well as provide golden opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives.[4]

Now think of the extent of details and the surge of data and information provided nowadays through the evolution in technologies and the internet. With the enlarge in storage competence and methods of data collection, huge amounts of data have become easily available. Every second, more and more data is being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data.

The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such utter amounts of big data need to be probably analyzed, and pertaining information should be extracted.

## 1.1 Business analytics

BA is often referred to as technologies, techniques, processes and methods to analyze data. It also incorporate people who turn data into insights to help companies understand their business superior and to support decision-making describe the term as “standard to communicate the full life cycle of magnify data-driven business decision making”. Moreover, they differentiate BA to business intelligence, as the idea of looking beyond the data level “to link a set of illustrative variables to a business response or outcome not only advanced reporting. Whereas business intelligence is explained as accessing data in a data warehouse by business users or applications to do reporting, querying or predictive analytics in an enterprise. The process of decision-making needs all aspects of the data. Therefore, BA includes prescriptive, descriptive, as well as predictive analytics and carry off techniques from statistics, machine learning, data mining and modeling.[3]

## 2. Big Data

The term “Big Data” refers to any set of data that is so large or so composite that conventional applications are not adequate to process them. The term also refers to the tools and technologies used to handle “Big Data”. Examples of Big Data include the amount of data shared in the internet everyday, YouTube videos viewed, twitter feeds and mobile phone location data. In the recent years, data produced by learning environments have also started to get big enough lift up the need for Big Data technologies and tools to handle them.[1]

### 2.1 CHALLENGES IN HANDLING BIG DATA

Several challenges need to be addressed while handling Big Data. Those challenges include

#### 2.1.1 Storage

While the common capacity of hard disks nowadays is in the range of terabytes, the amount of data generated through internet everyday is in the order of exabytes. Though the data generated in education is not as large as all the data generated through internet, it is large enough, and would get much larger in future. The traditional RDBMS tools will be unable to store or process such Big Data. To overcome this challenge, databases that don't use traditional SQL based queries are used. Compression technology is used to compress the data at rest and in memory.

### 2.1.2 Analysis:

As data generated to several types of online learning sites differ in structure and the size of the data is also huge, analysis of the data may devour a lot of time and resources. To overcome this, scaled out architectures are used to deal with the data in a distributed manner. Data are split into smaller pieces and processed in a vast number of computers accessible throughout the network and the processed data is aggregated.

### 2.1.3 Reporting:

Traditional reports necessitate display of statistical data in the form of numbers. When large amount of data are involved, traditional reports become difficult to interpret by human beings. In those cases the reports need to be appear in a form that can be easily acknowledge by looking into them.

### Enter Big Data: The Four Vs

One can separate big data and “regular-sized” data based on the presence of a set of characteristics commonly referred to as the four Vs: volume, variety, velocity, and veracity.[2]

#### Volume

The U.S. Library of Congress, which archives both digital and offline content, has collected hundreds of terabytes of data . Interestingly, the average company in 15 of 17 industry sectors in the United States has more data stored than the Library of Congress , which underscores the fact that big data is pervasive across industries including finance, manufacturing, retail, health, security, technology, and sports. For a detailed discussion of various applications domains for big data.

#### Variety

Organizations are now dealing with structured, semi-structured, and unstructured data from in and outside the enterprise . The variety includes traditional transactional data, user-generated text, images, and videos, social network data, sensor-based data, Web and mobile clickstreams, and spatial-temporal data. Effectively leveraging the variety of available data presents both opportunities and challenges.

#### Velocity

The speed of data creation is a hallmark of big data. For instance, Wal-Mart collects over 2.5 petabytes of customer transaction data every hour . With respect to unstructured data, over one billion new tweets occur every three days, and five billion search queries occur daily . Such information has important implications for “real-time” predictive analytics in various application areas, ranging from finance to health . Simply put, analyzing “data in motion” presents new challenges because the desired

patterns and insights are moving targets, which is not the case for static data.

#### Veracity

The credibility and reliability of different data sources vary. For instance, social media is plagued with spam, and Web spam accounts for over 20 percent of all content on the World Wide Web . Similarly, clickstreams from website and mobile traffic are highly susceptible to noise. Furthermore, deriving deep semantic knowledge from text remains challenging in many situations despite significant advances in natural language processing.

### 3. LIMITATIONS, FUTURE RESEARCH, CONCLUSION

The literature review might constitute a limitation of this paper. Only the databases AIS eLibrary, ProQuest and IEEE Xplore have been used for the literature search. Hence, other relevant papers related to BD and BA concepts and their impacts on companies’ competitive advantages could be mislaid. The selection of pertinence might have been affected by subjective interpretations as well. Since the literature only barely provides impacts on companies’ competitive advantages, it can be just assumed that the clustered concepts have the same or similar impacts. Furthermore, this paper only gives an idea what kind of impacts on companies’ competitive advantages can be perceive. No exact numbers of the amount were scrutinize for this paper; being a possibility for future research. We can conclude, that many companies like Capgemini or IBM performed studies about business analytics and big data and their sway on the companies’ competitive advantage in relation to decision - making processes, productivity, or revolution.[2]

However, a precise understanding of analogous costs and benefits, as well as a precise understanding of the impact on companies’ competitive advantage, was not establish, while impacts such as reduction of costs, improved decision-making, expand customer satisfaction, or operating logicity are known. Hence, our findings indicate that there is not “the concept” for big data and business analytics. Therefore, this paper is a first step towards a comprehensive understanding pointing out the relevant antecedents clustered in a general and technical overview of BIG DATA.

### REFERENCES

- [1] Katrina Sin and Loganathan Muthu (2015) Big Data in Education  
Data Minig and Learning analytics .ISSN:2229-6956
- [2] Ahmed Abbasi and Suprateek Sarkar ,Roger H.L.chiang Big Data Research in  
Informatic System:Toward an Inclusive  
research agenda ISSN:1536-9323

- [3] Fabiane Fernando ,Tobias Engel Big Data and Business Analytic Concepts.
- [4] Naga Elgendy and Ahmed Elragal (2014) Big Data analysis and its concepts.
- [5] Ritu Agarwal and Vasant Dhar (2014) Big Data ,Data science and analysis
- [6] The opportunity and challenge for Informatics system research :ISSN :1526-5536